September 23, 2008

Final Hotel Model Replication

This document contains the filter and variable calculation information necessary to replicate the final Hotel model. The final regression model uses data from the CBECS 2003 survey. ***This document is for internal reference purposes only.***

Notes on Variable Calculation

The final model requires a variety of calculations on the CBECS data in order to apply filters and format the variables for analysis. The following notes detail these calculations.

- Energy Calculations
  - Recodes: Observations that report "no use" for a given fuel type are not asked for a quantity. Therefore, for these observations it is appropriate to recode the system missing values to zero.
  - Source Energy: The following factors are applied to convert site energy values (in kBtu) in into Source Energy (in kBtu)
    - Electricity = 3.34
    - Natural Gas = 1.047
    - Fuel Oil = 1.01
    - Propane = 1.01
    - District Heat = 1.45 (if STUSED8=1)
    - District Heat = 1.35 (if only Hot Water is used, STUSED8=2)
  - Propane: The amount of propane must be estimated from the category selected (PRAMTC8). This will involve a "maximum" estimation for the purpose of applying filters and an "actual" estimation for use in the rating model. For the maximum, the high end of the propane category is used. For the actual, the middle of the category is used. These values are summarized in **Table 1**, below.

Table 1
Propane Estimation Guidelines

PRAMTC8	Maximum Estimation (filter)	Actual Estimation (model)
1: less than 100	99	50
2: 100 to 499	499	300
3: 500 to 999	999	750
4,7,8,9: 1000 or higher, or unknown	Not Included	Not Included

  - In order to convert the amount of propane in gallons or pounds into a Site energy value in kBtu, the following conversions are applied:
    - 1 Gallon of Propane = 92 kBtu
    - 1 Gallon of Propane = 4.5 Pounds

- Variable Recodes – In order to use the CBECS data in the regression, a few variables are re-coded, as follows:
    - If a respondent has selected "No" for cooling (COOL8 =2), then the percentage cooled (COOLP8) may be recoded from system missing into a value of zero.
    - If a respondent has selected "No" for heating (HT18 = 2 *and* HT28 =2), then the percentage heated (HEATP8) may be recoded from system missing into a value of zero.
    - The variable FDRM8 is recoded such that a response of "No" is represented by "0" rather than "2".
    - If the respondent does not have one of the types of commercial refrigeration (i.e. open cases, closed cases, or walk-in cases), then there will be system missing for the count of that case type. In these cases it is appropriate to re-code the missing counts to values of zero.

- General Calculations – In the final regression model, most of the variables have been computed to express the physical relationships (e.g. densities). Specific calculations are as follows:
    - Natural Log of Worker Density: LNWkrDen = LN(1000*NWker8/SqFt8)
        - All densities are computed on a per 1,000 square foot basis.
    - Heating Degree Days times Percent Heated: HDDxPH = HDD658*HeatP8/100
        - The percent is reported in the survey as an integer, but ultimately used in the model as a fraction.
    - Cooling Degree Days times Percent Cooled: CDDxPC = CDD658*CoolP8/100
    - Commercial Refrigeration Density: CommDen = 1000*(RFGWIN8+RFGOPN8+RFGCLN8)/SqFt8
    - Room Density: RoomDen = 1000*LODGRM8/SqFt8

- Variable Centering – For the purpose of the model, all of the variables are centered. Centering creates a situation where the intercept is equal to the average source energy intensity and the coefficients are used to adjust a building based on its deviation from the average value of each operating characteristic.
    - Centered Variable = Observation Value – Weighted Mean for that Variable
    - Centered LnWkrDen = LnWkrDen – Weighted Average LnWkrDen
    - Note that the weighted averages are computed across the filtered data set
    - Note that yes/no variables (i.e. FDRM8 are ***not*** centered). This allows for the interpretation of the intercept as the average without FDRM; the coefficient is the added allowance associated with FDRM.

Filters & Variable Weights

Four types of filters are applied to define the peer group for comparison and to overcome any technical limitations in the data: Building Type Filters, EPA Program Filters, Data Limitation Filters, and Analytical Filters. A complete description of each of these categories is provided in Section V of the general technical description document: *Energy Performance Ratings – Technical Methodology*. **Table 2** presents a summary of each filter applied in the development

of the Hotel model, the rationale behind the filter, and the resulting number of observations after the filter is applied. After all filters are applied, the remaining data set has 142 observations.

The reasons for applying filters on the use and quantity of propane are worthy of additional discussion. In CBECS, major fuel use is reported in exact quantities of consumption. However, if a building uses propane, the amount of propane is reported according to the variable PRAMT8, which uses ranges rather than exact quantities (e.g. less than 100 gallons, 100 to 500 gallons, etc). Therefore, the quantity must be estimated within the range. To limit error associated with this estimation, EPA applies two limits to the propane quantity.

1. The quantity of propane expressed by PRAMT8 must be 1000 gallons or smaller.
2. The value of propane cannot account for more than 10% of the total source energy use. Because the exact quantity of propane is not reported, this cap ensures that the quantity of propane entered will not introduce undue error into the calculation of total energy consumption. In order to apply this 10% limitation, the value at the high end of the propane category is employed (e.g. for the category of less than 100, a value of 99 is used). If the 10% cap is not exceeded, then EPA will use the value at the middle of the range to calculate total energy use (e.g. for the category of less than 100, a value of 50 is used).

Table 2 Summary of Filters		
Condition for *Including* an Observation in the Analysis	**Rationale**	**Number Remaining**
PBAPLUS8 = 38 or 39	Building Filter – CBECS defines building types according to the variable "PBAPLUS8." Hotels are coded as PBAPLUS=38 and Motels are coded as PBAPLUS=39. Both of these types are included in the analysis.	195
Must have at least 1 room	EPA Program Filter – Baseline condition for being a full time Hotel	195
Must operate for 168 hours per week	EPA Program Filter – Baseline condition for being a full time Hotel	192
Must have at least 1 worker	EPA Program Filter – Baseline condition for being a full time Hotel	190
Must operate for at least 10 months of the year	EPA Program Filter – Baseline condition for being a full time Hotel	182
A single activity must characterize greater than 50% of the floor space[1]	EPA Program Filter – In order to be considered part of the Hotel peer group, more than 50% of the building must be Hotel	180
If propane is used, the amount category (PRAMTC8) must equal 1, 2, or 3	Data Limitation – Cannot estimate propane use if it is "greater than 1000" or unknown	161
If propane is used, the maximum estimated propane amount must be 10% or less of the total source energy	Data Limitation – Estimation of propane cannot introduce too much error into the energy use value	159
Cannot use chilled water	Data Limitation – CBECS does not report chilled water consumption	157
Must have square foot >= 5,000	Analytical Limitation – Analysis could not model behavior for these smaller buildings	149
Must have EUI <=600 kBtu/ft^2	Analytical Limitation – Values determined to be statistical outliers	147
Must have EUI >=10 kBtu/ft^2	Analytical Limitation – Values determined to be statistical outliers	146
Must have fewer than 5 rooms per 1,000 square foot	Analytical Limitation – Values determined to be statistical outliers	143
Must have fewer than 0.5 commercial refrigeration units per 1,000 square foot[2]	Analytical Limitation – Values determined to be statistical outliers	142

[1] This filter is applied by a set of screens. If the variable ONEACT8=1, then one activity occupies 75% or more of the building. If the variable ONEAC8T=2, then the activities in the building are defined by ACT1, ACT2, and ACT3. One of these activities must be coded as lodging (PBAX=21), with a corresponding percent (ACT1PCT8, ACTPCT8, ACT3PCT8) that is greater than 50.

[2] This filter is based on the variable Commercial Refrigeration Density: CommDen. This variable is defined above under the *General Calculations* section.

Descriptive Statistics

Key descriptive statistics can be computed across this final filtered dataset (142 observations). The variable weights are applied for all calculations, and in the regression (ADJWT8). **Table 3** presents basic descriptive statistics for variables in the model. A table of descriptive statistics for a broader set of variables is attached.

Table 3
Summary of Descriptive Statistics for Hotels

Variable Name	Number of Buildings	Minimum	Maximum	Mean	Std. Deviation
SourceEUI	53865	49.0781	544.3123	182.5338	93.3999
RoomDen	53865	0.5195	4.2373	1.9511	0.8193
LNWkrDen	53865	-3.2452	1.0079	-1.3949	0.7346
FDRM8	53865	0.0000	1.0000	0.2056	0.4042
RfgCommDen	53865	0.0000	0.3125	0.0227	0.0510
HDDxPH	53865	31.9000	9928.0000	4120.1693	2624.5146
CDDxPC	53865	0.0000	4871.0000	1223.9798	915.7863

Regression Output

The final regression is a weighted ordinary least squares regression across the filtered data set of 142 observations. The dependent variable is SourceEUI. Each independent variable is centered relative to the weighted mean, as described above. The **Tables 4 through 6** present the final model outputs from SPSS. Residual plots are attached. As shown in Table 4, the model has an R^2 value of 0.367. This value is computed in units of Source Energy Intensity (kBtu/ft^2), the dependent unit of analysis. The R^2 value indicates that the model explains 36.7% of the variation in Source EUI among hotels. When EUI is the dependent variable, the explanatory power of square foot is not included in the R^2 value, thus this value appears artificially low. Re-computing the R^2 value in units of *source energy* demonstrates that the model actually explains 87.3% of the variation of source energy in hotels. This is an excellent result for a statistically based energy model. There is an attached table presenting the R^2 calculations in units of Source Energy and LN(Source Energy).

Table 4: Model Summary Output from SPSS

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.606(a)	.367	.339	1484.43973

a Predictors: (Constant), C_CDDxPC, C_RfgCommDen, C_RoomDen, FDRM8, C_LNWkrDen, C_HDDxPH

Table 5: ANOVA(b,c) Output from SPSS

Model		Sum of Squares	df	Mean Square	F	Sig.
	Regression	172406044.578	6	28734340.763	13.040	.000(a)
	Residual	297480777.263	135	2203561.313		
	Total	469886821.841	141			

a Predictors: (Constant), C_CDDxPC, C_RfgCommDen, C_RoomDen, FDRM8, C_LNWkrDen, C_HDDxPH
b Dependent Variable: SrcEUI
c Weighted Least Squares Regression - Weighted by ADJWT8

Table 6: Coefficients(a,b) Output from SPSS

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	169.140	7.457		22.682	.000
	C_RoomDen	33.218	9.330	.291	3.560	.001
	C_LNWkrDen	20.805	10.381	.164	2.004	.047
	FDRM8	65.137	18.644	.282	3.494	.001
	C_RfgCommDen	249.778	147.203	.136	1.697	.092
	C_HDDxPH	0.010737	.003	.302	3.653	.000
	C_CDDxPC	0.01685	.008	.165	1.988	.049

a Dependent Variable: SrcEUI
b Weighted Least Squares Regression - Weighted by ADJWT8

Lookup Table

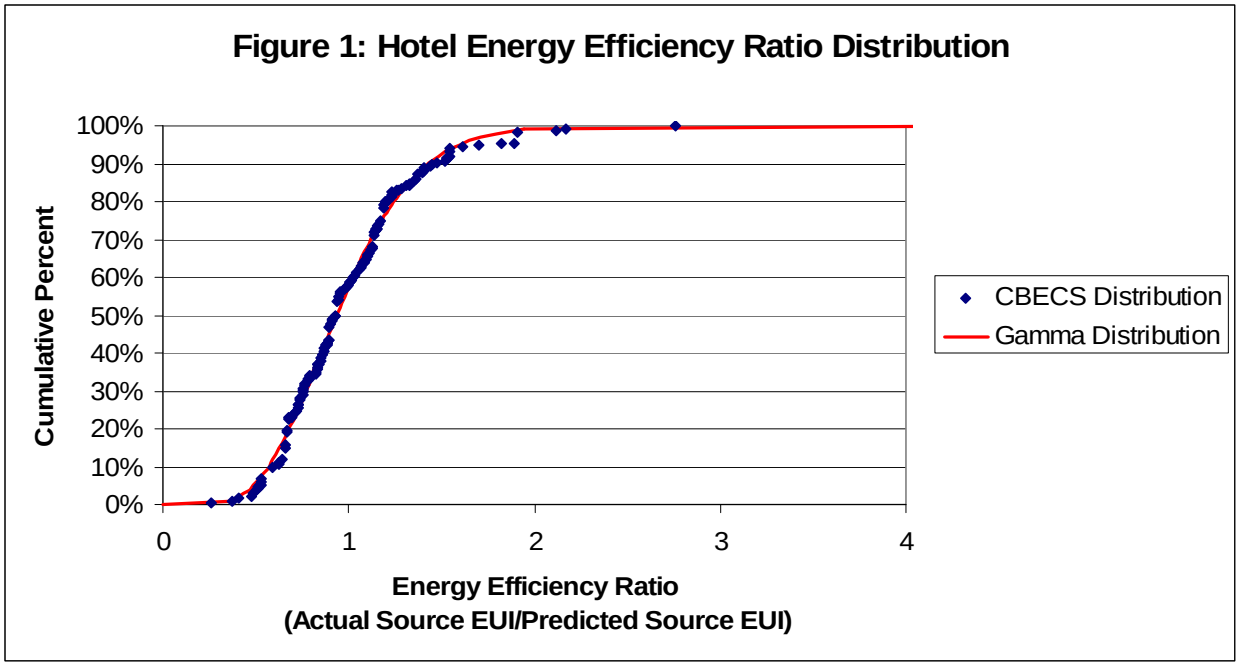
The model is used to generate a scoring lookup table according to the following steps:

1. The model is used to generate a predicted Source EUI for each observation.
2. An energy efficiency ratio is calculated for each observation as follows:
   Efficiency Ratio = Actual Source EUI/Predicted Source EUI
3. The weighted cumulative percent at each observation's ratio is computed. For a given observation:
   a. Cumulative percent = (sum of weights for all observations with ratios less than or equal to the observation) / (sum of all weights)
4. The energy efficiency ratios are then fitted to a gamma distribution, where the parameters alpha and beta are estimated in order to minimize the sum of square differences between the actual and the gamma values for the cumulative distribution at each observation[3].
5. For the hotel model, the final gamma distribution is computed to have the following values:
   a. Alpha (shape) = 8.0805

[3] This fit can be achieved using MS Excel's Solver function. The actual cumulative distribution (weighted) is computed for each observation. Then, arbitrary alpha and beta values are assigned and used to compute a gamma distribution value for each observation. The solver can be set to minimize the sum of the differences between the actual and gamma distribution values across all observations, by changing both alpha and beta. The resulting alpha and beta are the final shape and scale values, respectively.

b. Beta (scale) = 0.1205
c. Sum of squared differences: 0.0485

6. The validity of this fit can be verified graphically, see **Figure 1**.
7. The Alpha and Beta value are then use to find the ratio that corresponds to each point (0.01 to 1.0) on the distribution. Note that the rounded values presented in Step 5 are used to generate this fit. The fit can be performed as an inverse gamma function within Microsoft Excel. The lookup table is constructed such that each point represents 1 percent; that is, a rating of 99 corresponds to the ratio at a cumulative percent of 0.01. The final lookup table is attached.



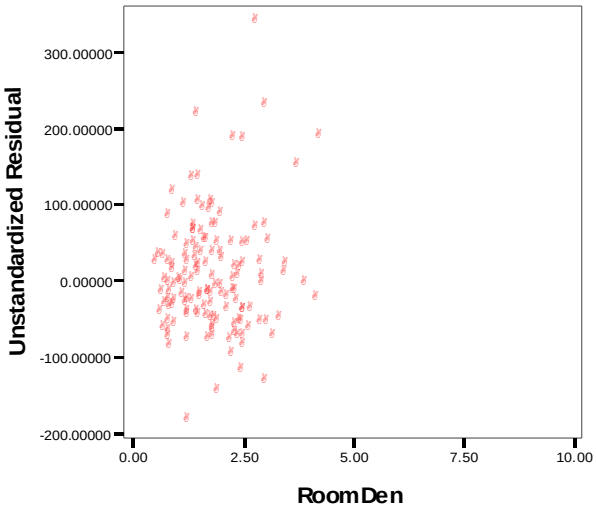
Attachment – Complete Descriptive Statistics

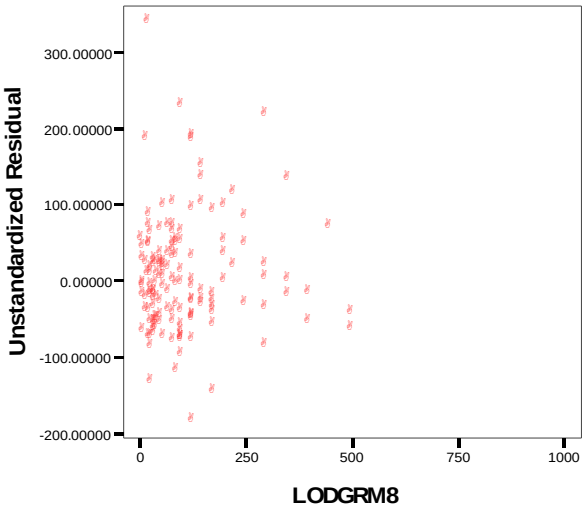
Table 7 Expanded Set of Descriptive Statistics for Hotels					
	N	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
SrcEUI	53865	49.0781	544.3123	182.5338	93.3999
Source	53865	641422.3560	104269685.1840	7863620.9518	14229950.2877
LNSource	53865	13.3714	18.4625	15.1019	1.1384
RoomDen	53865	0.5195	4.2373	1.9511	0.8193
LNWkrDen	53865	-3.2452	1.0079	-1.3949	0.7346
FDRM8	53865	0.0000	1.0000	0.2056	0.4042
RfgCommDen	53865	0.0000	0.3125	0.0227	0.0510
HDDxPH	53865	31.9000	9928.0000	4120.1693	2624.5146
CDDxPC	53865	0.0000	4871.0000	1223.9798	915.7863
LODGRM8	53865	8.0000	500.0000	62.0066	69.6067
NWKER8	53865	1.0000	400.0000	14.6960	36.8402
HEATP8	53865	10.0000	100.0000	89.7003	23.5767
COOLP8	53865	0.0000	100.0000	87.4715	26.4004
WkrDen	53865	0.0390	2.7397	0.3199	0.2515

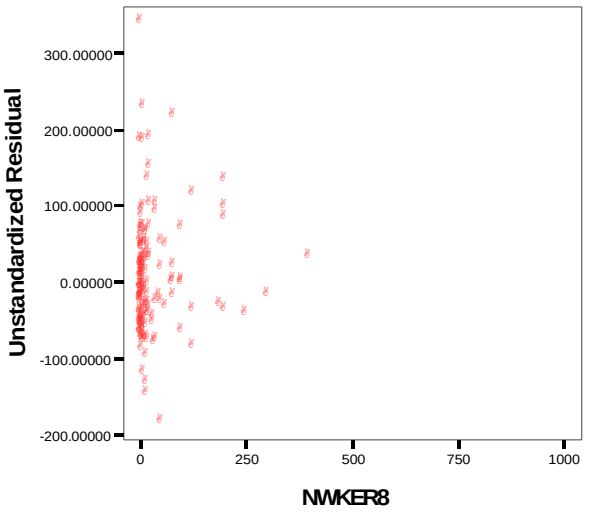
Attachment – R^2 Calculations

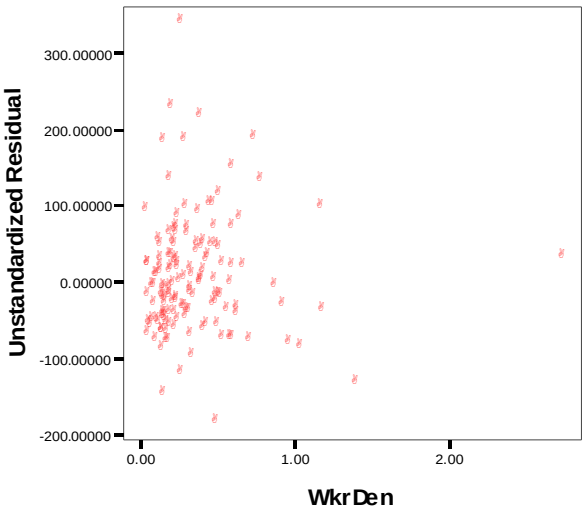
Table 8 R^2 Calculations	
Units of Source Energy	
Residual Variation	1.08811E+16
Total Variation	8.58043E+16
Source R^2	0.873187096
Units of LN(Source	
Residual Variation	18.84486633
Total Variation	312.4388957
LNSource R^2	0.939684634

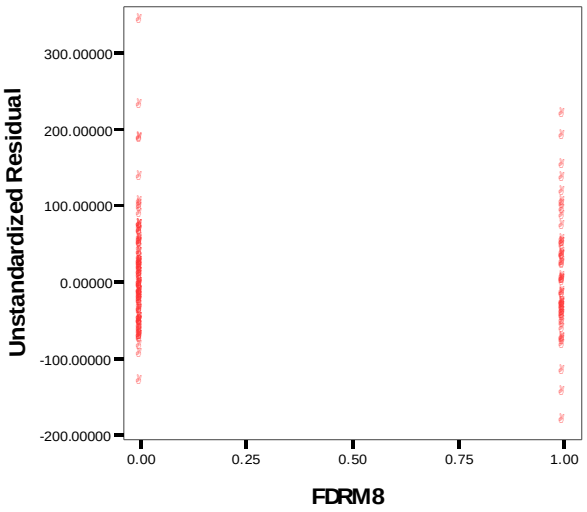
Attachment – Residual Plots

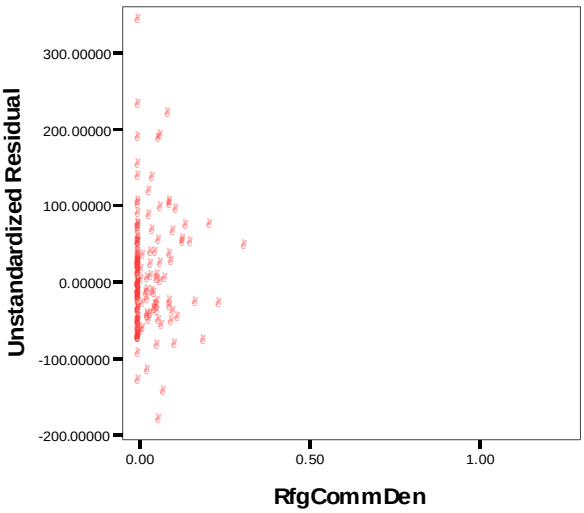


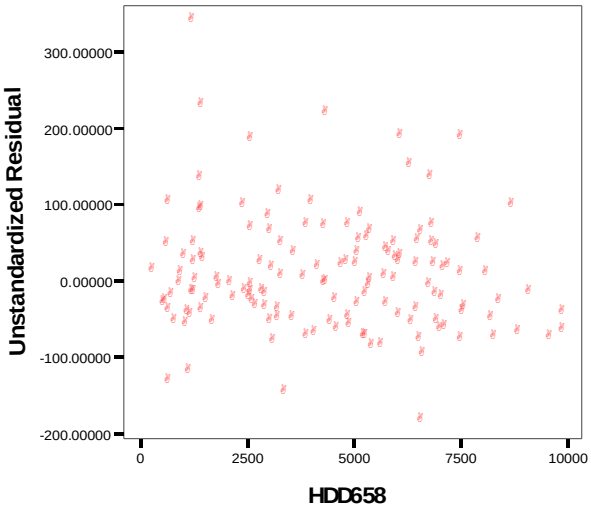


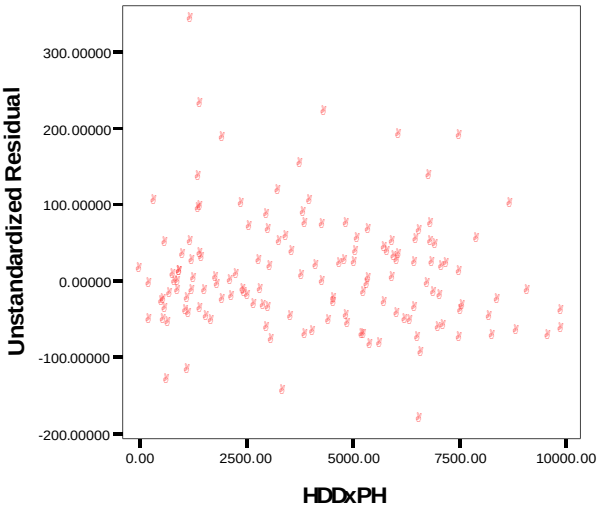


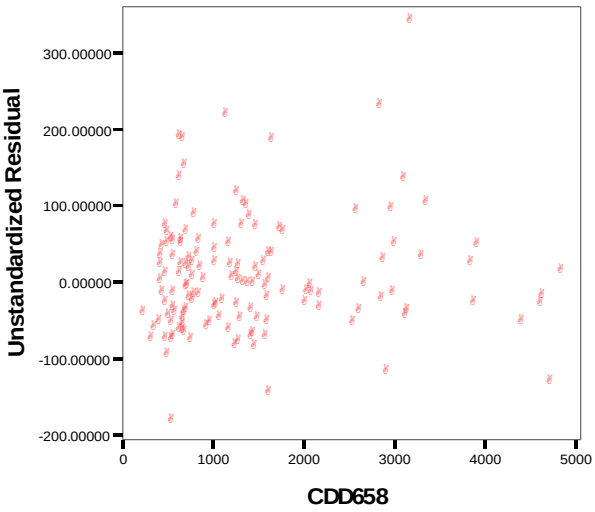


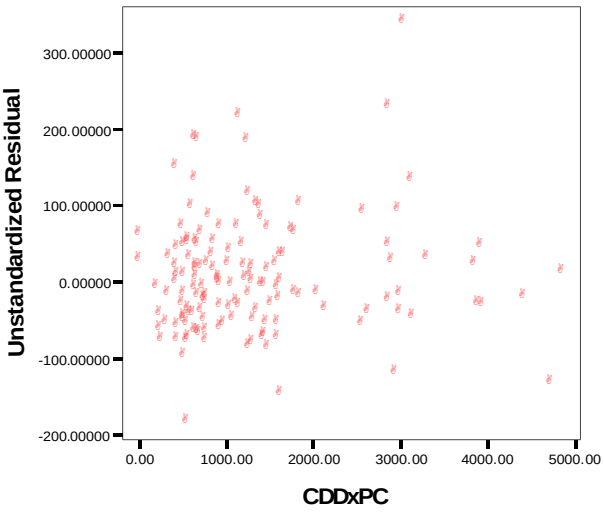

Unstandardized Residual
300.00000
200.00000
100.00000
0.00000
-100.00000
-200.00000
0.00
0.25
0.50
0.75
1.00
FDRM8


Unstandardized Residual
300.00000
200.00000
100.00000
0.00000
-100.00000
-200.00000
0.00
0.50
1.00
RfgCommDen


Unstandardized Residual
300.00000
200.00000
100.00000
0.00000
-100.00000
-200.00000
0
2500
5000
7500
10000
HDD658


Unstandardized Residual
300.00000
200.00000
100.00000
0.00000
-100.00000
-200.00000
0.00
2500.00
5000.00
7500.00
10000.00
HDDxPH


Unstandardized Residual
300.00000
200.00000
100.00000
0.00000
-100.00000
-200.00000
0
1000
2000
3000
4000
5000
CDD658


Unstandardized Residual
300.00000
200.00000
100.00000
0.00000
-100.00000
-200.00000
0.00
1000.00
2000.00
3000.00
4000.00
5000.00
CDDxPC

Attachment – Final Lookup Table

Final Hotel Lookup Table						
Rating	**Cumulative Percent**	**Ratio**		**Rating**	**Cumulative Percent**	**Ratio**
100	0%	0		50	50%	0.933842
99	1%	0.355913		49	51%	0.942246
98	2%	0.404678		48	52%	0.950706
97	3%	0.437998		47	53%	0.959227
96	4%	0.464319		46	54%	0.967815
95	5%	0.486541		45	55%	0.976477
94	6%	0.506038		44	56%	0.985217
93	7%	0.523581		43	57%	0.994044
92	8%	0.539647		42	58%	1.002963
91	9%	0.554555		41	59%	1.011982
90	10%	0.568531		40	60%	1.021108
89	11%	0.581738		39	61%	1.030349
88	12%	0.594302		38	62%	1.039714
87	13%	0.606319		37	63%	1.049211
86	14%	0.617867		36	64%	1.05885
85	15%	0.629008		35	65%	1.068641
84	16%	0.639794		34	66%	1.078595
83	17%	0.650266		33	67%	1.088724
82	18%	0.660462		32	68%	1.099041
81	19%	0.670412		31	69%	1.109559
80	20%	0.680142		30	70%	1.120294
79	21%	0.689677		29	71%	1.131261
78	22%	0.699036		28	72%	1.14248
77	23%	0.708237		27	73%	1.15397
76	24%	0.717297		26	74%	1.165752
75	25%	0.726229		25	75%	1.177852
74	26%	0.735047		24	76%	1.190296
73	27%	0.743762		23	77%	1.203115
72	28%	0.752386		22	78%	1.216344
71	29%	0.760927		21	79%	1.230021
70	30%	0.769396		20	80%	1.244191
69	31%	0.777801		19	81%	1.258904
68	32%	0.786149		18	82%	1.274221
67	33%	0.794448		17	83%	1.290209
66	34%	0.802705		16	84%	1.30695
65	35%	0.810926		15	85%	1.32454
64	36%	0.819118		14	86%	1.343095
63	37%	0.827287		13	87%	1.362753
62	38%	0.835439		12	88%	1.38369
61	39%	0.843579		11	89%	1.406123
60	40%	0.851713		10	90%	1.430329
59	41%	0.859846		9	91%	1.456672
58	42%	0.867983		8	92%	1.485643
57	43%	0.876129		7	93%	1.517923
56	44%	0.88429		6	94%	1.554503
55	45%	0.892471		5	95%	1.596907
54	46%	0.900676		4	96%	1.647663
53	47%	0.90891		3	97%	1.711455
52	48%	0.917179		2	98%	1.798656
51	49%	0.925488		1	99%	1.941713